

A background image of a sunset or sunrise over a dark, silhouetted landscape. The sky is filled with clouds, and the sun is a bright orange circle near the horizon, casting a warm glow across the scene.

Developing a Metadata Template for CDC

**The Efforts of the Metadata and Data Quality
Subgroup to Capture Data about Data**

A background image of a sunset or sunrise. The sun is a bright, glowing orb positioned slightly to the left of the center, partially obscured by a dark, silhouetted treeline. The sky is filled with soft, orange and yellow clouds, creating a warm, atmospheric glow.

A Presentation by Jeff Patridge

**On Behalf of the Metadata and Data Quality Subgroup
of the Environmental Public Health Tracking
Standards and Network Development Workgroup**

A bright sun is positioned in the upper left quadrant of the image, casting a strong glow and creating a lens flare effect across the sky. The sky is a deep, clear blue, and several fluffy white clouds are scattered throughout, particularly concentrated in the lower half of the frame. The word "Background" is centered in the middle of the image in a bold, black, serif font.

Background

What is Metadata?

Metadata are “Data About Data”. They help a person to locate and understand data by describing the content, quality, condition, and other characteristics of the data.

Metadata Reveals...

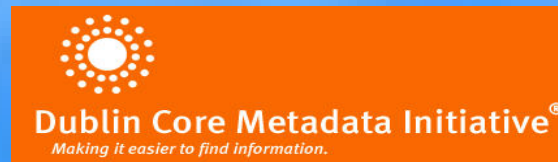
- **What method(s) were used to collect data.**
- **When were data collected?**
- **How were data processed?**
- **When were data last updated?**
- **Are there any data gaps?**

Why is Metadata Important?

- **Protects investment in data**
- **Helps users to understand data**
- **Allows for users to discover the existence of data**
- **Limits liability**
- **Can reduce staff workload (once created)**

Commonly Used Metadata Standards

- **Federal Geographic Data Committee (FGDC) Content Standards for Digital Spatial Metadata**
- **Dublin Core**
- **ISO 11179**
- **ISO 19115**
- **ISO 19139** (*currently undergoing review*)



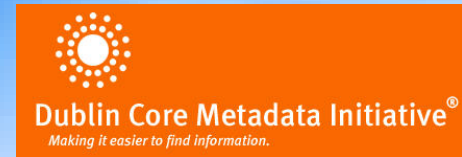
Federal Geographic Data Committee (FGDC)



- Standard created for documenting geospatial datasets.
- Presidential Executive Order 12906 establishes that geospatial datasets created by Federal agencies must include FGDC-compliant metadata.
- Basic elements include:

Dataset Title	Purpose	Access Constraints
Contact Info	Citation	Time Period of Dataset
Status	Spatial Domain	Keywords
Attributes	Distribution	Metadata Reference

Dublin Core



Dublin Core is a higher level metadata standard. It consists of 16 elements and several element refiners. The elements are:

Coverage

Description

Type

Relation

Source

Subject

Title

Audience

Contributor

Creator

Publisher

Rights

Date

Format

Identifier

Language

International Organization for Standardization

ISO 11179



- **Specifies a basic set of data element characteristics necessary to share data.**
- **Metadata about data elements is stored in a data element registry.**
- **Basic attributes of data elements include:**

Name	Classification Scheme	Data Type
Identifier	Keywords	Maximum Size
Version	Related Data Reference	Minimum Size
Context	Type of Relationship	Permissible Values

International Organization for Standardization

ISO 19115



- **ISO 19115 incorporates the FGDC standard.**
- **Allows for the documenting of both geographic and non-geographic data.**
- **Will be superceded in the United States by ISO 19139**

International Organization for Standardization

ISO 19139



- **Based on ISO 19115**
- **Extensible Markup Language (XML) model**
- **Currently undergoing review / revision**
- **Technical specification designation Winter 2004-2005**

Importance of Metadata to the EPHT Network

Metadata and Data Quality Subgroup formed as part of Standards and Network Development Workgroup to:

- **Develop a metadata template using a controlled vocabulary of EPHT Network datasets that will identify a core set of information that is needed to adequately document a dataset and its limitations for potential users.**
- **Develop a means to describe data using common ways to document datasets to facilitate data searches.**

Actions Taken by Subgroup

- **Talked with individuals involved with Public Health Information Network (PHIN) and the EPA Exchange Network**

Actions Taken by Subgroup

- **Talked with individuals involved with Public Health Information Network (PHIN) and the EPA Exchange Network**
- **Reviewed data inventories of the grantees to determine common elements**

Actions Taken by Subgroup

- **Talked with individuals involved with Public Health Information Network (PHIN) and the EPA Exchange Network**
- **Reviewed data inventories of the grantees to determine common elements**
- **Evaluated common elements against the currently accepted metadata standards**

Results of Element Mapping

- **Dublin Core too general to meet the requirements for describing an electronic dataset**

Results of Element Mapping

- Dublin Core too general to meet the requirements for describing an electronic dataset
- **FGDC and ISO 19115 addressed most of the identified elements**

Results of Element Mapping

- Dublin Core too general to meet the requirements for describing an electronic dataset
- FGDC and ISO 19115 addressed most of the identified elements
- **FGDC recommended as the standard for the Network, until superceded by ISO standard.**

A bright sun is positioned in the upper left quadrant of the image, casting a strong glow and creating a lens flare effect across the sky. The sky is a deep, clear blue, and several fluffy white clouds are scattered throughout, particularly concentrated in the lower half of the frame. The word "Testing" is centered in the middle of the image, overlaid on the clouds.

Testing

Swat Team Mission

- **Develop a test for the FGDC Content Standard**
- **Test the usefulness of the freeware tool Tkme**
- **Gather feedback from test participants**
- **Make recommendations to Metadata Subgroup members.**



Test Participants

- **Illinois**
- **Maryland**
- **Missouri**
- **New Mexico**
- **New York City**
- **Oklahoma**
- **Oregon**



Recommendations

- **Determine a core set of FGDC elements that will constitute the bare minimum data elements that will be required to put data on the EPHTN. Additional elements could be added if applicable.**
- **Develop a tool that allows the easy creation of metadata. The tool should be user-friendly and should have built in constraints.**

A bright sun is positioned in the upper left quadrant of the image, casting a strong glow and creating a lens flare effect across the sky. The sky is a deep, clear blue, and it is populated with numerous white, fluffy clouds of varying sizes and densities. The clouds are more concentrated in the lower half of the image, with some smaller, wispy clouds scattered higher up. The overall scene is bright and clear, suggesting a sunny day.

Core Metadata Template

Element	Field	Definition
Section 1: IDENTIFICATION		
Citation	Citation (1.1)	Information to be used to reference the data set.
Originator	Originator (8.1)	The name of an organization or individual that developed the data set. If the name of editors or compilers are provided, the name must be followed by "(ed.)" or "(comp.)" respectively.
Publication Date	Publication_Date (8.2)	The date when the data set is published or otherwise made available for release.
Title	Title (8.4)	The name by which the data set is known.
URL	On-line_Linkage (URL) (8.10)	The name of an online computer resource that contains the data set. Entries should follow the Uniform Resource Locator convention of the Internet.
Description	Description (1.2)	A characterization of the data set, including its intended use and limitations.
Abstract	Abstract (1.2.1)	A brief narrative summary of the data set.
Purpose	Purpose (1.2.2)	A summary of the intentions with which the data
Supplemental Info	Supplemental_Info (1.2.3)	Other descriptive information about the data set.
Time Period of Content	Time_period_of_content (1.3)	Time period(s) for which the data set corresponds to the currentness reference. (Can be a single date, multiple dates, or a range of dates.).
Calendar Date	Calendar_Date (9.1.1)	The year (optionally month or month and day).
Currentness	Currentness_Reference (1.3.1)	The basis on which the time period of content information is determined.
Status	Status (1.4)	The state of and maintenance information for the data set.
Progress	Progress (1.4.1)	The state of a data set ("complete, in work, planned").
Maintenance and Update Frequency	Maintenance_and_Update_Frequency (1.4.2)	The frequency that changes are made to the data set after the initial data set is completed.
Spatial Domain	Spatial_Domain (1.5)	The geographic area covered by the data set.
West Bounding Coordinate	West_Bounding_Coordinate (1.5.1.1)	Western-most coordinate of the limit of coverage expressed in longitude.
East Bounding Coordinate	East_Bounding_Coordinate (1.5.1.2)	Eastern-most coordinate of the limit of coverage expressed in longitude.
North Bounding Coordinate	North_Bounding_Coordinate (1.5.1.3)	Northern-most coordinate of the limit of coverage expressed in latitude.
South Bounding Coordinate	South_Bounding_Coordinate (1.5.1.4)	Southern-most coordinate of the limit of coverage expressed in latitude.

Element	Field	Definition
Keywords	Keywords (1.6)	Words or phrases summarizing an aspect of the data set.
Theme Keyword Thesaurus	Theme_Keyword_Thesaurus (1.6.1.1)	Reference to a formally registered thesaurus or a similar authoritative source of theme keywords.
Theme Keyword	Theme_Keyword (1.6.1.2)	Common-use word or phrase used to describe the subject of the data set.
Place Keyword Thesaurus	Place_Keyword_Thesaurus (1.6.2.1)	Reference to a formally registered thesaurus or a similar authoritative source of place keywords.
Place Keyword	Place_Keyword (1.6.2.2)	The geographic name of a location covered by a data set. (Includes city, county, state, state acronym, regional descriptions and references).
Access Constraints	Access_Constraints (1.7)	Restrictions and legal prerequisites for accessing the data set. These include any access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the data set.
Use Constraints	Use_Constraints (1.8)	Restrictions and legal prerequisites for using the data set after access is granted. These include any use constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on using the data set.
Point of Contact	Point_of_Contact (1.9)	Contact information for an organization that is knowledgeable about the data set. Format found in Section 10.
Native Data Set Environment	Native_Data_Set_Environment (1.13)	A description of the data set, including the name of the software, computer operating system, file name, and data set size.

Section 5: ENTITY AND ATTRIBUTES

Overview Description	Overview_Description (5.2)	Description of the entities, attributes, attribute values, and related characteristics encoded.
Entity and Attribute Overview	Entity_and_Attribute_Overview (5.2.1)	Detailed summary of the information contained in a data set.
Entity and Attribute Detail Citation	Entity_and_Attribute_Detail_Citation (5.2.2)	Reference used to the complete description of the entity types, attributes, and attribute values for the data set.

Element	Field	Definition
Section 7: METADATA REFERENCE		
Metadata Date	Metadata_Date (7.1)	The date that the metadata were created or last updated.
Metadata Contact	Metadata_Contact (7.4)	The party responsible for the metadata information. Format found in Section 10.
Metadata Standard Name	Metadata_Standard_Name (7.5)	The name of the metadata standard used to document the data set.
Metadata Access Constraints	Metadata_Standard_Version (7.6)	Restrictions and legal prerequisites for accessing the metadata. These include any access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the metadata.
Metadata Use Constraints	Metadata_Access_Constraints (7.8)	Restrictions and legal prerequisites for using the metadata after access is granted. These include any metadata use constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on using the metadata.
Section 10: CONTACT INFORMATION		
Contact Information	Contact_Information (10.0)	This section provides a means of identifying individuals and organizations, and is used by other sections of the metadata standard. This section is never used alone.
Contact Organization	Contact_Organization (10.1.2)	The name of the organization.
Contact Address	Contact_Address (10.4)	The address for the organization.
Address Type	Address_Type (10.4.1)	Address type ("mailing address"; "physical address"; "mailing and physical address").
Address	Address (10.4.2)	Contact address for organization.
City	City (10.4.3)	Contact address city.
State or Province	State_or_Province (10.4.4)	Contact address state or province.
Postal Code	Postal_Code (10.4.5)	Contact address Zip or postal code.
Country	Country (10.4.6)	Contact address country.
Contact Telephone Number	Contact_Voice_Telephone (10.5)	The telephone number by which individuals can speak to the organization.
Contact Fax Number	Contact_Facsimile_Telephone (10.7)	The telephone number of a facsimile machine of the organization.
Contact E-mail Address	Contact_Electronic_Mail_Address (10.8)	The address of the electronic mailbox of the organization.
<p>Template is based on a subset of the Federal Geographic Data Committee's "Content Standard for Digital Geospatial Metadata" Version 2 - 1998. (FGDC-STD-001 June 1998). The complete FGDC standard can be viewed at www.fgdc.gov/metadata/contstan.html</p> <p>*Areas in gray are descriptive headings. Data entry elements are in white.</p>		



Metadata Tool Requirements Gathering

Team Mission

- **Create a prototype tool to spark discussion**
- **Gather tool requirements**
- **Package those requirements for use by a developer**



Next Steps...

- **Complete compiling the metadata creation tool requirements**
- **Begin compiling requirements for the creation of a metadata registry**
- **Continue to promote the creation and use of metadata**



Thanks for Listening

For further information on anything seen in this presentation please contact:

Jeff Patridge, GIS Analyst

Missouri Department of Health and Senior Services

Office of Surveillance

930 Wildwood Drive

Jefferson City, MO 65109-0570

Email: patrij1@dhss.mo.gov